REPORT DOCUMENTATION PAGE

Form Approved OMB NO. 0704-0188

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED (From - To)
14-06-2017	Final Report	5-Apr-2013 - 4-Sep-2016

4. TITLE AND SUBTITLE
Final Report: Scalable Matrix Algorithms for Interactive Analytics of Very Large Informatics Graphs

5a. CONTRACT NUMBER
W911NF-13-1-0072

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
611102

6. AUTHORS
Michael Saunders, Michael Mahoney

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES AND ADDRESSES
Stanford University
3160 Porter Drive
Suite 100
Stanford, CA 94304 -8445

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS (ES)
U.S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSOR/MONITOR'S ACRONYM(S)
ARO

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
63856-NS.12

12. DISTRIBUTION AVAILIBILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES
The views, opinions and/or findings contained in this report are those of the author(s) and should not contrued as an official Department of the Army position, policy or decision, unless so designated by other documentation.

14. ABSTRACT
The research developed, implemented, and evaluated traditional matrix and graph algorithms at large scale to allow an analyst with domain insight to explore more interactively the properties of large, e.g., consisting of millions or billions of nodes, social and information networks. Depending on the situation, these larger networks may not fit on a single machine. Although we considered traditional matrix and graph algorithms, e.g., regression and low-rank matrix approximation, we took a nontraditional approach: we extended recent work on randomized matrix algorithms in order to implement them in parallel and distributed environments that are appropriate for networks

15. SUBJECT TERMS
randomized linear algebra, social networks

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	15. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Michael Saunders
UU	UU	UU	UU		19b. TELEPHONE NUMBER 650-723-1875

Report Title

Final Report: Scalable Matrix Algorithms for Interactive Analytics of Very Large Informatics Graphs

ABSTRACT

The research developed, implemented, and evaluated traditional matrix and graph algorithms at large scale to allow an analyst with domain insight to explore more interactively the properties of large, e.g., consisting of millions or billions of nodes, social and information networks. Depending on the situation, these larger networks may not fit on a single machine. Although we considered traditional matrix and graph algorithms, e.g., regression and low-rank matrix approximation, we took a nontraditional approach: we extended recent work on randomized matrix algorithms in order to implement them in parallel and distributed environments that are appropriate for networks that are so large that they may be difficult to store on a single machine. By using several complementary methods, this will allow the downstream analyst to test hypotheses as to what properties of the output are realistic properties of the data, given what he or she knows about the sociological mechanisms that generated the data, and what are artifacts of the formalization and algorithms used to explore the data. Currently, this is common for small network analysis, and a result of this research is to make this type of interactive analytics easier for much larger networks.

Enter List of papers submitted or published that acknowledge ARO support from the start of the project to the date of this printing. List the papers, including journal references, in the following categories:

(a) Papers published in peer-reviewed journals (N/A for none)

Received		Paper
06/14/2017	3	Lucas Jeub, Prakash Balachandran, Mason Proter, Peter Mucha, Michael Mahoney. Think Locally, Act Locally: The Detection of Small, Medium-Sized, and Large Communities in Large Networks, ArXiv, PRE, (03 2014): 0. doi: 329,771.00
06/14/2017	1	Jiyan Yang, Xiangrui Meng, Michael Mahoney. Quantile Regression for Large-scale Applications, ArXiv e-prints / JMLR / SISC, (05 2013): 0. doi: 300,682.00
06/14/2017	4	Aaron Adcock, Blair Sullivan, Michael Mahoney. Tree decompositions and social graphs,, arXiv / Internet Mathematics, ( ): 0. doi: 358,519.00
06/14/2017	5	Jiyan Yang, Yin Lam, Chris Re, Michael Mahoney. Weighted SGD for Lp Regression with Randomized Preconditioning, arXiv / SODA / JMLR, ( ): 0. doi: 358,520.00
06/14/2017	8	Aaron Adcock, Blair Sullivan, Michael Mahoney. Teeelike structure in large social and information networks, ICDM, ( ): . doi: 1,045,356.00
06/14/2017	9	David Gleich, Michael Mahoney. Using local spectral methods to robustify graph-based learning algorithms,, Proc of the ACM KDD, ( ): . doi: 1,045,357.00
06/14/2017	10	Lucas Jeub, Michael Mahoney, Mason Porter, Peter Mucha. arXiv / Network Science, A Local Perspective on Community Structure in Multilayer Networks, ( ): . doi: 1,045,358.00
06/14/2017	11	Nate Veldt, David Gleich, Michael Mahoney. A Simple and Strongly-Local Flow-Based Method for Cut Improvement, arXiv / ICML, ( ): . doi: 1,045,359.00
TOTAL:	**8**	

Number of Papers published in peer-reviewed journals:

(b) Papers published in non-peer-reviewed journals (N/A for none)

Received	Paper

TOTAL:

Number of Papers published in non peer-reviewed journals:

(c) Presentations

Number of Presentations: 0.00

Non Peer-Reviewed Conference Proceeding publications (other than abstracts):

Received	Paper

TOTAL:

Number of Non Peer-Reviewed Conference Proceeding publications (other than abstracts):

Peer-Reviewed Conference Proceeding publications (other than abstracts):

Received		Paper
08/10/2015	7.00	David F. Gleich , Michael W. Mahoney. Using Local Spectral Methods to Robustify Graph-Based Learning Algorithms, KDD15. 10-AUG-15, . : ,
08/11/2014	2.00	Aaron Adcock , Blari Sullivan , Michael Mahoney. Tree-like Structure in Large Social and Information Networks, Proc. ICDM. 07-DEC-13, . : ,

TOTAL: **2**

Number of Peer-Reviewed Conference Proceeding publications (other than abstracts):

(d) Manuscripts

Received Paper

TOTAL:

Number of Manuscripts:

Books

Received Book

TOTAL:

Received Book Chapter

08/10/2015 6.00 David F. Gleich, Michael W. Mahoney. Mining Large Graphs, CRC Handbook: CRC Press, (12 2015)

TOTAL: **1**

Patents Submitted

Patents Awarded

Awards

NA

Graduate Students

NAME	PERCENT_SUPPORTED	DISCIPLINE
Jiyan Yang	50	ICME
Aaron Adcock	50	ICME
FTE Equivalent:	**1.00**	
Total Number:	**2**	

Names of Post Doctorates

NAME	PERCENT_SUPPORTED
FTE Equivalent:	
Total Number:	

Names of Faculty Supported

NAME	PERCENT_SUPPORTED
FTE Equivalent:	
Total Number:	

Names of Under Graduate students supported

NAME	PERCENT_SUPPORTED
FTE Equivalent:	
Total Number:	

Student Metrics

This section only applies to graduating undergraduates supported by this agreement in this reporting period

The number of undergraduates funded by this agreement who graduated during this period: 0.00

The number of undergraduates funded by this agreement who graduated during this period with a degree in science, mathematics, engineering, or technology fields:..... 0.00

The number of undergraduates funded by your agreement who graduated during this period and will continue to pursue a graduate or Ph.D. degree in science, mathematics, engineering, or technology fields:..... 0.00

Number of graduating undergraduates who achieved a 3.5 GPA to 4.0 (4.0 max scale):..... 0.00

Number of graduating undergraduates funded by a DoD funded Center of Excellence grant for Education, Research and Engineering:..... 0.00

The number of undergraduates funded by your agreement who graduated during this period and intend to work for the Department of Defense 0.00

The number of undergraduates funded by your agreement who graduated during this period and will receive scholarships or fellowships for further studies in science, mathematics, engineering or technology fields:..... 0.00

Names of Personnel receiving masters degrees

NAME
Total Number:

Names of personnel receiving PHDs

NAME
Jiyan Yang
Aaron Adcock
Total Number: 2

Names of other research staff

NAME	PERCENT_SUPPORTED
FTE Equivalent:	
Total Number:	

Sub Contractors (DD882)

Inventions (DD882)

Scientific Progress

Scientific publications that arose from this project include the following:

A Local Perspective on Community Structure in Multilayer Networks,
L. G. S. Jeub, M. W. Mahoney, P. J. Mucha, and M. A. Porter,
Technical Report, Preprint: arXiv:1510.05185 (2015) (arXiv),
Network Science, 5(2): 144-163, 2017 (pdf).

A Simple and Strongly-Local Flow-Based Method for Cut Improvement,
N. Veldt, D. F. Gleich, and M. W. Mahoney,
Technical Report, Preprint: arXiv:1605.08490 (2016) (arXiv),
Proc. of the 33rd ICML Conference 1938-1947 (2016) (pdf), (supp).

Using local spectral methods to robustify graph-based learning algorithms,
D. F. Gleich and M. W. Mahoney,
Proc. of the 21st Annual SIGKDD, (2015) (pdf) (code).

Weighted SGD for Lp Regression with Randomized Preconditioning,
J. Yang, Y.-L. Chow, C. Re, and M. W. Mahoney,
Technical Report, Preprint: arXiv:1502.03571 (2015) (arXiv),
Proc. of the 27-th Annual SODA, 558-569 (2016) (pdf),
Accepted for publication, J. Machine Learning Research.

Tree decompositions and social graphs,
A. B. Adcock, B. D. Sullivan, and M. W. Mahoney,
Technical Report, Preprint: arXiv:1411.1546 (2014) (arXiv), (code).
Internet Mathematics, 12(5), 315-361 (2016) (pdf).

Think Locally, Act Locally: The Detection of Small, Medium-Sized, and Large Communities in Large Networks,
L. G. S. Jeub, P. Balachandran, M. A. Porter, P. J. Mucha, and M. W. Mahoney,
Technical Report, Preprint: arXiv:1403.3795 (2014) (arXiv), (code, code),
Physical Review E, 91, 012821 (2015) (pdf).

Tree-like Structure in Large Social and Information Networks,
A. B. Adcock, B. D. Sullivan, and M. W. Mahoney,
Proc. of the 2013 IEEE ICDM, 1-10 (2013) (pdf).

Quantile Regression for Large-scale Applications,
J. Yang, X. Meng, and M. W. Mahoney,
Technical Report, Preprint: arXiv:1305.0087 (2013) (arXiv), (code),
Proc. of the 30th ICML Conference, JMLR W&CP 28(3): 881-887 (2013) (pdf),
SIAM J. Scientific Computing, 36(5), S78-S110 (2014) (pdf).

Here are links to associated publicly-available code for different parts of this project:

https://github.com/LJeub/LocalCommunities
https://www.cs.purdue.edu/homes/dgleich/codes/robust-diffusions/
https://github.com/chocjy/randomized-LS-solvers
https://github.com/aadcock/tree-like-network-lib
https://github.com/chocjy/randomized-quantile-regression-solvers

Technology Transfer

NA